

AI QUICKSTART

Deploy Biological Foundation Models in 6 Weeks

Models like **Tahoe-x1** are now publicly available—but deploying them for real scientific use still requires infrastructure, integration & reproducible workflows.

DataXight AI QuickStart is a fixed-scope, 6-week engagement that delivers:

- A working AI inference environment on your infrastructure
- Executable notebooks tailored to your workflows
- Canary tests and benchmarked evaluation
- Deployment tooling and documentation for full handoff

The service enables **life sciences organizations** to evaluate and operationalize state-of-the-art **biological AI models** without committing to a **proprietary platform** or **long-term managed service**.

 **TAHOE-x1**

 **ESM-2**

 **Geneformer**

Evo-2

scGPT

AlphaFold2

WHAT YOU GET

A production-ready AI deployment foundation — a scalable stack designed for your team to expand long after our engagement concludes. The primary AI inference environment, production-ready notebooks & validated evaluation results.

{ 1 } MODEL SELECTION AND DEPLOYMENT

- **Scoping conversation** in week 1 against your **use case**, data, and downstream workflows
- **Open-weight model** pulled from canonical sources with **version pinning & provenance**
- **Containerized inference** environment on your cloud (DNAexus, AWS, GCP) or on-prem cluster
- **GPU configuration** tuned to your workload; deployment scripts & container definitions included in the handoff

Current model menu: Tahoe-x1 (perturbation and cell-state work), Geneformer (cell-type annotation), ESM-2 (protein representation), Evo-2 (genomic sequence modeling), scGPT (single-cell representation and cross-cell inference), AlphaFold2 (protein structure prediction).

{ 2 } EXECUTABLE NOTEBOOKS

- Working **end-to-end** notebooks parameterized for your data
- Workflows the deployed model supports: embedding generation, cell-state queries, perturbation lookups, or sequence and structure workflows as applicable
- Optional **template notebooks** for evidence integration against Open Targets, GTEx, and pathway databases.

{ 3 } CANARY TESTS AND EVALUATION

- **Evaluation report** reproducing the published benchmarks on your deployment
- **Canary notebooks** that validate endpoint outputs against reference inputs on a schedule
- Optional benchmarking against internal reference data you provide
- Documentation covering model provenance, environment versions, and known failure modes

{ 4 } FULL OWNERSHIP, NO VENDOR LOCK-IN

AI Quickstart delivers a reproducible **foundation-model** environment on your infrastructure—complete with executable notebooks, benchmarked evaluation, and **full ownership** of all deployment artifacts and workflows.

FROM DATA TO INSIGHT
POWERED BY SOFTWARE
INNOVATION

Contact us today to discuss your project at solutions@dataught.com

AI QUICKSTART

Deploy Biological Foundation Modes in 6 Weeks

Models like **Tahoe-x1** are now publicly available—but deploying them for real scientific use still requires infrastructure, integration & reproducible workflows.

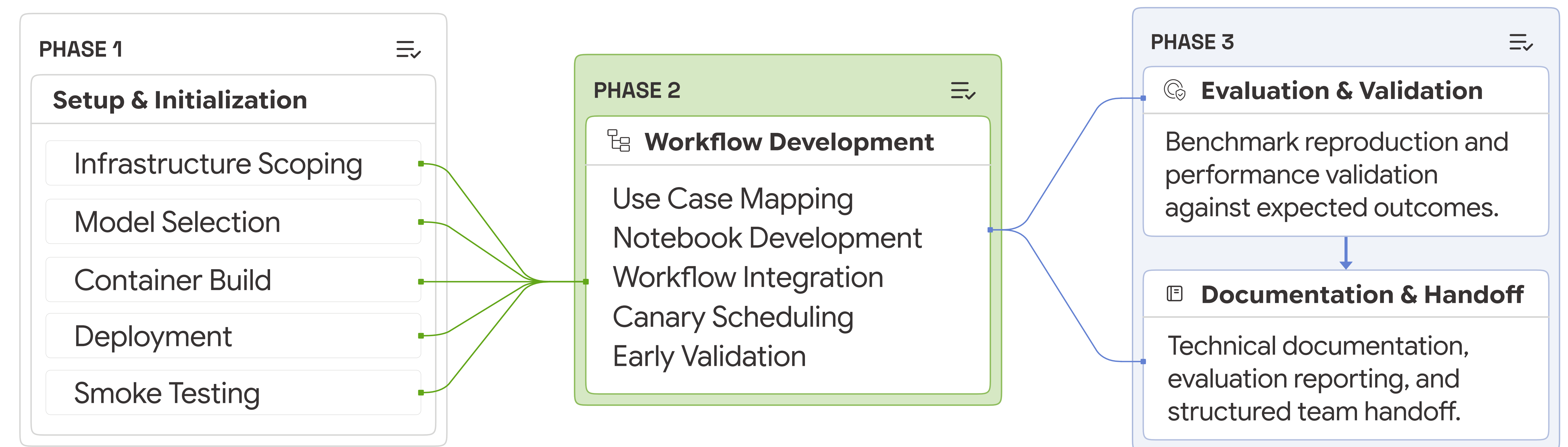
HOW IT WORKS

{ 1 } INPUT ALIGNMENT:

You provide **system access (cloud or on-prem)** with GPU availability, and a defined use case.

{ 2 } DATA XIGHT TIMELINE:

Fixed Scope. 6-week engagement.



Weeks 1–2

Working inference environment on your infrastructure; models selected, deployed and validated.

Weeks 3–4

Reproducible notebooks tailored to your biology; initial workflows running on schedule.

Weeks 5–6

Benchmark reproduction, evaluation report, documentation, working session with your team.

{ 3 } FULL STACK OWNERSHIP:

You own the full stack at the end: infrastructure definitions, deployment scripts, model artifacts, notebooks, evaluation report, and documentation — in a git repository you control. **No ongoing license, no vendor lock-in.**

WHO IT'S FOR

- **Pharma and biotech R&D teams** under executive mandate to put foundation models into internal use
- **Research computing** and **computational biology groups** that want foundation-model inference available on their own systems, fully self-managed, with optional support from DataXight if needed
- Teams that need a working environment on a **discovery-phase** timeline, not a multi-quarter platform build

OPTIONAL SERVICES

Analysis engagements (target prioritization, disease signature reports, candidate ranking) are scoped separately, as are **fine-tuning** and **continued pretraining**. **AI QuickStart** delivers the environment in which those become tractable.

PREREQUISITES

- **Infrastructure with GPU access** (A100 or H100 preferred; A10G or L4 workable for smaller workloads)
- **Container registry** and access controls sufficient for deployment into a **sandbox project, compartment, or namespace**
- **A defined use case** — the specific biology the notebooks should demonstrate

FROM DATA TO INSIGHT
POWERED BY SOFTWARE
INNOVATION

Contact us today to discuss your project at solutions@dataxight.com